A Novel Method for Network Intrusion Detection System Using Data Mining Methodologies

Shaik Faamida[1], Sudhakar Putheti[2]
[1]P.G. Scholar, [2]Professor

[1, 2] Department of CSE, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India

Abstract - Intrusion detection is one of the big security issues in the cyber world today. A large number of techniques based on machine learning approaches have been developed. However, all forms of intrusions are not very effective. A detailed study and review of various machine learning techniques has been undertaken in this paper to determine the causes of problems associated with specific machine learning techniques for disruptive activities. The description of attacks and mapping of attack characteristics is given for each attack. Issues related to the identification of low-frequency attacks with data from network attacks are also addressed and feasible approaches for improvement are suggested. Machine learning methods for the identification of various types of attacks were analyzed and compared. Limitations for each group are also discussed. Numerous machine learning data mining methods have also been included in the report. Finally, potential directions for attack detection using machine learning techniques are given.

Keywords: Anomaly Detection, Intrusion Detection System, Network Security, Principal Component Analysis

I. INTRODUCTION

The use of the internet and internet-based software has grown exponentially in the present computing environment, with a proportionate rise in intrusions as cyber attacks. The control of emerging ways of disruption is a challenging challenge for the government and a global problem.

The main purpose of intrusion detection is to investigate resources, identify suspicious activities and abuses. In 1980, Anderson adapted the model from 1980[3] by offering different practices to improve consumer security auditing and supervision. Between 1984 and 1986 the initial IDS, P Neumann and D Denning developed known as Intrusion Detection Expert System (IDES). IDES was initially trained to detect malicious activity using a rule-based approach and then transformed into Next Generation IDS (NIDES). The University of California and the United States in 1988. Project programs funded by government such as Haystack (US Air Force). Research was accomplished by contrasting tests with established behavior trends, developing host-driven behavioral pattern matching mechanisms and integrating them in the disseminated setting in 1990, introducing Davis Todd to Network-Based Intrusion Detection (NIDS), and contributing to DIDS and NSM (Network Security Monitoring) as well as to digital IDS (Computer Mi) in the early 90's. ASM (Automated Safety Measurement System) was introduced in 1994.

Figure 1 demonstrates the general working process of IDS. The purpose and role of IDS is to prevent unauthorized safety attacks on machines and networks. The behavioral characteristics of phenomena have taken on a new face and the current IDS is difficult to identify. An intrusion can occur internally or externally in the computer world. For the current monitoring method and defense system, the changing existence of emerging security threats is difficult to evaluate. The view of statistics warns us that we are vulnerable to network irregularities in terms of security breaches. Another big problem is the study of intrusion footprints.

The basic and critical aspect of IDS in today's digital computing world is network security design. For IDS representation, the perception of an intrusion is important. The performance of IDS can be measured based on the honesty, confidentiality and simple usability of the system. If an action or occurrence violates the machine's confidentiality, it can be viewed as an intrusion. In general, these types of activities occur when an attacker attempts unauthorized access to the network or services. In certain cases, malicious software can be designed to cause intrusions that breach the legality of the device. Many external actions threaten to breach network or system accessibility. The conventional system architecture has firewalls which failed to detect malicious intrusions and cannot completely protect the systems.

The consequences of intrusions will deny legitimate users functionality and services cannot be used. To provide full network security, the new digital age requires advanced monitoring mechanisms and expanded IDS prevention schemes. IDS 'most important tasks are to analyze network data packets to detect malicious tasks, to track the flow of data streams and to restrict the successful mitigation of the system's operation. Figure 1 displays the basic functional and operational modules of generic IDS.

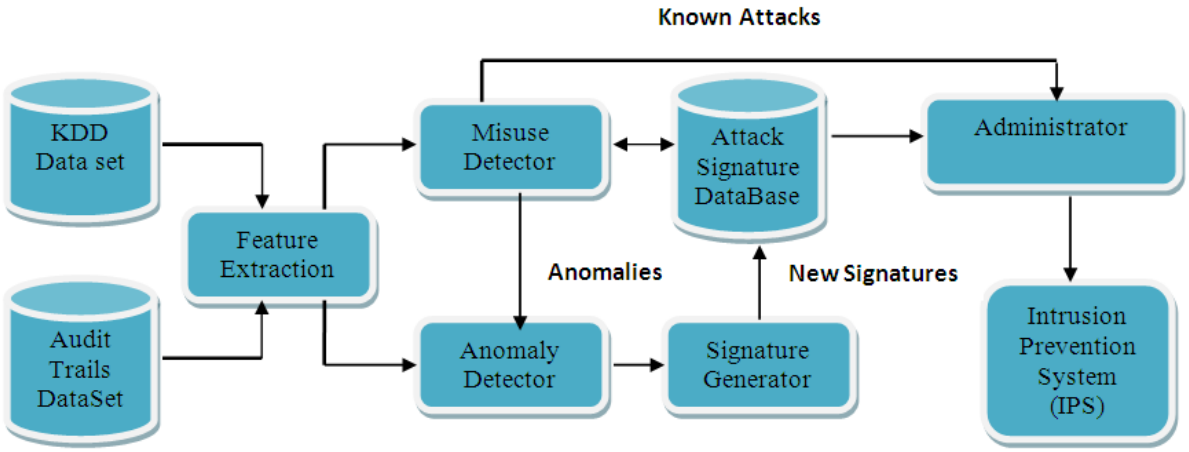

Figure 1. Generic View of Intrusion Detection

II. RELATED WORK

Abdulla et al., 2016[1] indicated that mobile agents were used to detect intrusions, and that flooding attacks such as DoS and DDoS attacks were caught in the cloud environment. The snort and suricata output is both implemented on different systems such as Linux, FreeBSD and ESXi, and findings were compared with Alhomoud et al. , 2011 [2].

Mainly Anderson has proposed the method of intrusion detection in 1980[3]. Buczak, 2015 [4] has indicated that learning machines and data mining techniques are not adequate to detect cyber traffic misuse and anomaly in real time. The suggested approach has shown less precision.

Abid et al., 2017[5] recommended performance on the Intel Berkeley real-life data base and used in WSNs to detect performance measurements such as DR, FAR, and accuracy. For through iterative operation, different numbers of test cases are taken into account. Snort has demonstrated better performance than Suricata in windows, related operating systems. The suggested approach is based on profiles in terms of predictive metrics for behavioral symptoms.

Bellovin, 1994[6] addressed firewall limitations and firewall security loopholes. The layered firewall architecture is introduced and possibilities and risks are discussed.

Blum et al., 1997[7] proposed that the algorithms of the selection attributes be used for the machine learning. The authors have experimented with fewer attributes on a broad dataset.

Carlos and Carlos, 2012[8] suggested an IDS solution that takes more time to prepare data and increases the network administrator load.

Chiu et al. proposed an unattended scenario to lower the false alarm rate in 2010[9]. They proposed a fast detection filtering system. In both semi-supervised and supervised methods, the suggested method uses the same features.

In order for the improvement of intruder detection using unlabeled data, Ching-Hao et al., 2009[10] recommended a co-training system, which showed a lower error rate compared to existing methodologies. Turner et al., 2016[11] tested various snort rules versions. Most rules are disabled in all snort versions of rules that trigger inadequate security protection. There is plenty of space to write more complex snort rules to increase health standards.

Das 2001[12] said a hybrid algorithm named BBHFS was used to improve the efficiency of the learning methods and the ID3 classification methodology used for the classification of data sets, which is a fairly low efficiency method compared to a vector machine. Dasgupta and Gonzalez suggested a multi-level risk control for both identified and unknown attacks in 2001[13]. We used a standard intrusion detection method focused on laws that cannot detect intrusions with high extreme values.

Denning proposed a system for identification and tracking of suspicious patterns in audit data for the prevention in security breaches in 1987 [14]. Hybrid PSO technology that can accommodate nominal attributes was used. The suggested approach with a basic rule set demonstrates enhanced accuracy.

Dickerson and Dickerson, 2000[15] recommended a fluoridated solution to the logic based engine of intrusion detection. They used small data sets for experimental purposes, which in effect allow the estimation of the feature subset to reduce the rate of accuracy. Divya and Lakra 2013[16] proposed a hybrid snort intrusion detection method that uses artificial intelligence to detect only a few types of

attacks, and the results obtained were not very accurate. Reddy et al., 2011[17] suggested various ways to construct an intrusion sensing device. They elaborated the data mining principles and introduced some advanced techniques for IDS design.

III. PROPOSED SYSTEM

The suggested solution consists of several steps including pre-processing of the attributes, removing redundant and unnecessary attributes and classifying the characteristic vector as normal or anomalous. This section explains the key components of the algorithm. The proposed model structure is defined in Figure 2.

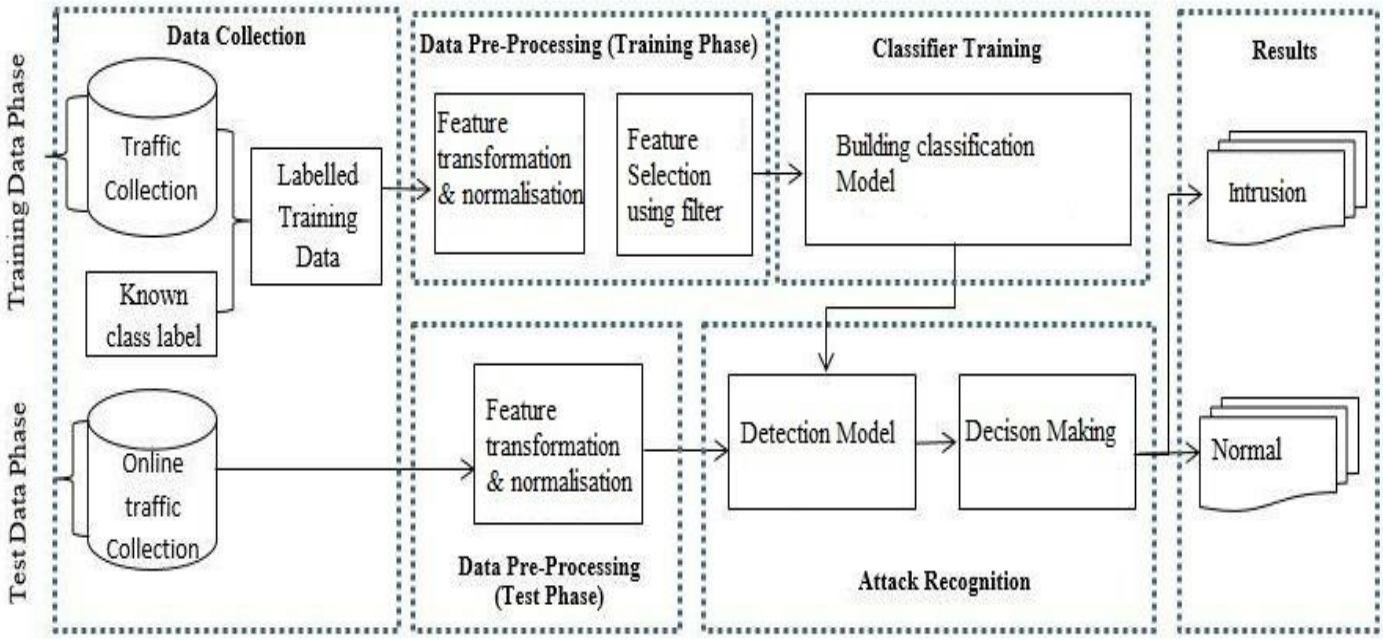

Figure 2. Proposed Architecture

3.1. Data Set Preparation- 10% updated KDD Cup99 dataset is used for data preparation. The dataset contains approximately 400,000 documents. These reports have to do with 23 different assaults. For this analysis a sample dataset size of 5857 records is chosen for iterative data samples consisting of different types of attacks.

3.2. Data Pre-processing – Data cleaning operations on the dataset are conducted to make the dataset transparent. WEKA is used to conduct data pre-processing operation with the aid of free open source data mining software. An unregulated RemoveUseless() filtering technique is used in the compilation phase to remove redundant attributes from 41 attributes.

3.3. Data Classification- Efficient kernel-based vector support method for classifying data. Algorithm features of support vector machines are used to identify the data collection.

3.4. Intrusion Detection Module: Free open source sniffing application used to detect intrusion. Snort is paired with the method for collecting network data packets for the WinPcap packet.

IV. RESULTS AND DISCUSSION

The classification results of the classifier proposed are contrasted with most rising current approaches. In the R tool setting the suggested classifier is implemented. The classifier suggested has more appropriate values than current methodologies. Table 1 displays the performance reference values.

Table 1. Results comparison

	Accuracy	**Sensitivity**	**Specificity**	**FAR**
Naïve Bayes	90.92	83.30	98.17	1.83
J48	95.03	91.32	98.57	1.43
Random Forest	95.73	92.68	98.63	1.37
SMO	94.55	91.25	97.70	2.30
C4.5	86.56	82	93.22	1.55
SVM	83.81	64.28	87.17	3.21
(ACO , C4.5)	90.35	86.13	96.15	0.79
(SVM, ACO)	88.39	70.86	89.76	3.46
(C4.5, PSO)	91.7	89.76	96.02	0.93
(SVM, PSO)	91.35	69.1	92.16	0.93
EDADT	97.75	89.71	97.51	0.21
EKBSVM	**99.81**	**99.9**	**99.62**	**0.07**

V. CONCLUSION

The success of the kernel-based SVM methodology has increased the accuracy of current methodologies. False alarm levels have been minimized; this reduced FAR would impact administrator workload reduction directly. The proposed method provided 14.24 percent sensitivity is increased by 14.96 percent over C4.5 compared to the SVM approach. At the other hand, the proposed system registered a sensitivity inclination of 11.45 percent over C4.5+ACO and its increased values over EDADT are 2.69 percent.

VI. REFERENCES

1. Abdelali, Saidi & Ali, Kartit & Bendriss, Elmehdi & El marraki, Mohamed. (2016). The functional of a mobile agent system to enhance DoS and DDoS detection in cloud. 11. 4615-4617.

2. Alhomoud M, Adeeb & Munir, Rashid & Pagna Diss, Jules & Awan, Irfan & Al-

3. Dhelaan, Abdullah. (2011). Performance Evaluation Study of Intrusion Detection Systems. Procedia CS. 5. 173-180. 10.1016/j.procs.2011.07.024.
4. Anderson P, James. (1980). Computer Security Threat Monitoring and Surveillance.
5. Anna, Buczak & Guven, Erhan. (2015). A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials. 18. 1-1. 10.1109/COMST.2015.2494502.
6. Aymen Abid et al.:, Outlier detection for wireless sensor networks using density-based clustering approach, IET Wireless. Sens. Syst., 2017, Vol. 7 Iss. 4, pp. 83-90, The Institution of Engineering and Technology 2017, ISSN 2043-6386
7. Bellovin, S.M. "Network Firewalls", IEEE Communications Magazine, Vol. 32, pp. 50- 57, 1994.
8. Blum, Avrim L. & Pat Langley (1997). Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1-2), 245–271
9. Catania Carlos A, Garino Carlos. Automatic network intrusion detection: current techniques and open issues. Elsevier Comput Electr Eng 2012; 38(5):1062–72.
10. Chien-Yi Chiu, Yuh-Jye Lee, Chien-Chung Chang. Semi-supervised learning for false alarm reduction. In: Industrial conference on data mining, no. 10; 2010. p. 595–605.
11. Ching-Hao, Hahn-Ming L, Devi P, Tsuhan C, Si-Yu H. Semi-supervised co-training and active learning based approach for multi-view intrusion detection. In: ACM symposium on applied computing, no. 9; 2009. p. 2042– 7.

12. Claude Turner et al. (2016). A Rule Status Monitoring Algorithm for Rule-Based Intrusion Detection and Prevention Systems, Complex Adaptive Systems, Conference Organized by Missouri University of Science and Technology 2016 - Los Angeles, CA, Procedia Computer Science 95 (2016) 361 – 368, 1877-0509, doi: 10.1016/j.procs.2016.09.346
13. Das, S. (2001). Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. Proc. 18thInt'l Conf. Machine Learning, 74-81
14. Dasgupta, D. and F. A. Gonzalez, "An intelligent decision support system for intrusion detection and response", In Proc. Of International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM-ACNS), St.Petersburg. Springer- , 21-23 May,2001.
15. Denning, D.E. "An Intrusion-Detection Model", in IEEE Transactions on Software Engineering, Vol.13, No. 2, pp. 222-232, 1987.
16. Dickerson, J. E. and J. A. Dickerson, "Fuzzy network profiling for intrusion detection", In Proc. of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta, pp. 301-306, 2000 North American Fuzzy Information
17. Divya and Surendra Lakra, "SNORT: A Hybrid intrusion detection system using artificial intelligence with a snort", International journal computer technology & application, Vol 4(3), 466-470, 2013.
18. E.Kesavulu Reddy, Member IAENG, V.Naveen Reddy, P.Govinda Rajulu. A Study of Intrusion Detection in Data Mining. Proceedings of the World Congress on Engineering 2011 Vol III WCE 2011, July 6 - 8, 2011, London, U.K.
19. Eskin, E., Arnold, A., Prerau, M., Portnoy, L., and Stolfo, S. J., A Geometric Framework for Unsupervised Anomaly Detection: Detecting Intrusions in Unlabeled Data, In D.Barbarà and S. Jajodia (eds.), Applications of Data Mining in Computer Security, Kluwer Academic Publishers, Boston, MA, 2002, pp. 78-99.
20. Fayyad, U. M., G. Piatetsky-Shapiro, and P. Smyth, "The KDD process for extracting useful Knowledge from volumes of data," Communications of the ACM 39 (11), November 1996, 2734.
21. G. J. Klir, "Fuzzy arithmetic with requisite constraints", Fuzzy Sets and Systems, 91:165175, 1997.